NMR Data Management

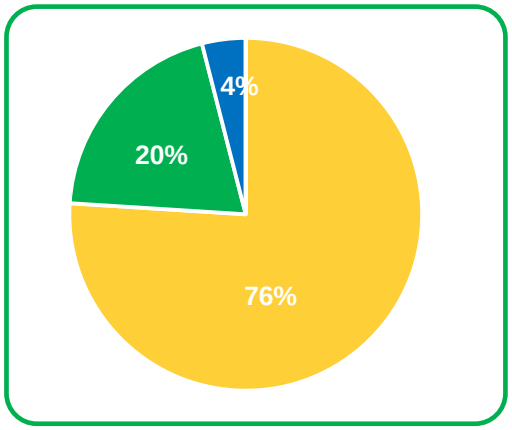
UKMRM Meeting 2016 - Oxford

Online survey – Academia vs Industry

Academia – 27 Responses (87 %)

Industry – 4 Responses (13 %)

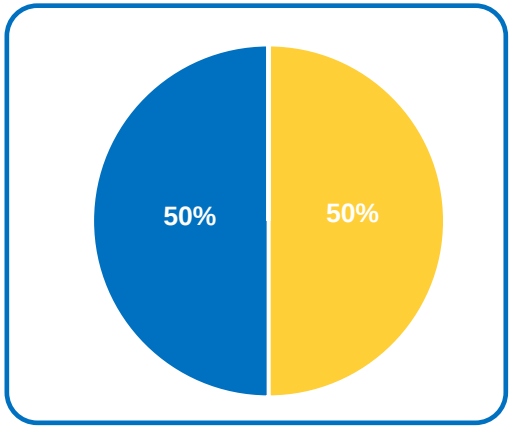
Does your NMR data store provide a search facility?



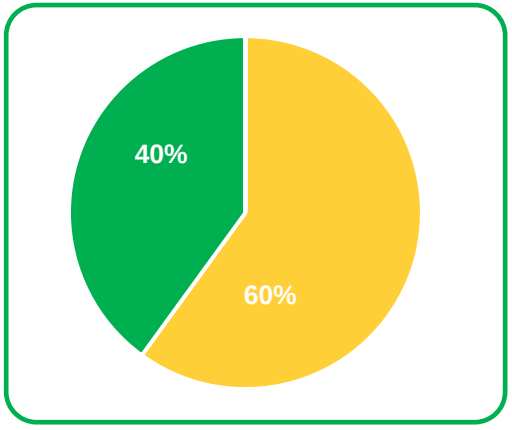
Yes. SQL database.

Yes. Other.

No



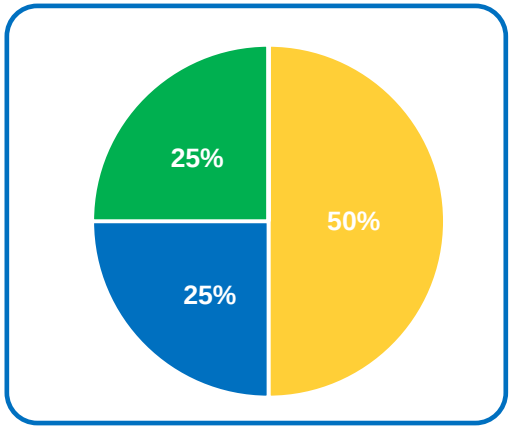
Does your institution use ELN (Electronic Lab Notebook)?



Yes. Compulsory

Yes. Some Groups.

No



There has been no driving force to improve scientific data management in academia

Funder's Open Access Data Policies

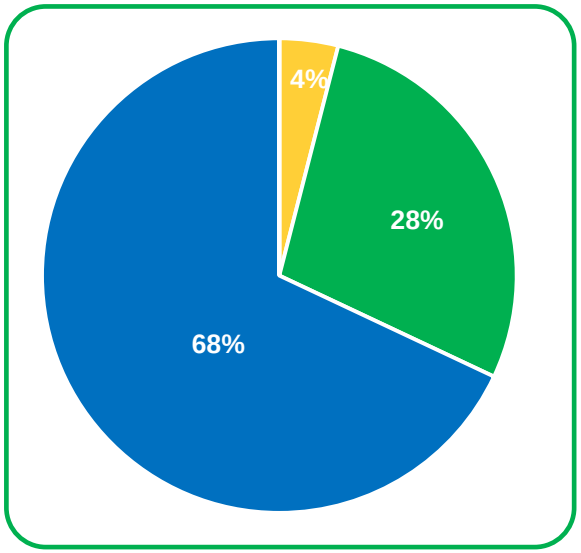
- All RCUK (EPSRC, STFC, BBSRC, etc.), EC (H2020), the Royal Society, the Wellcome Trust
- Mainly apply to published results

- **Data Management Plan (DMP):**
  - Describes the files that will be generated (formats, volume), how they will be stored, how compliance will be achieved, what will be shared; where data will be archived
- **Data:**
  - should be made publicly available
  - should be retained - at least *<10 years>* from last date of access
  - include a statement on how to access supporting data (where? On which conditions?)
- **Costs:**
  - in most cases they can be included in grant applications

- **FASTR** (Fair Access to Science and Technology Research Act of 2015)
  - Bill introduced to US senate last year
  - In summary - every federal agency that spends more than $100 million on grants for research would be required to adopt an open access policy.

Online survey – Open Access Data Policies

How do you deal with the Open Access Data Policies?



17	I leave this problem to PIs, institution management or IT services.
7	I pro-actively try to change the way of archiving the NMR data to accommodate the policy.
1	I have no idea what this is.
0	I have been forced to accommodate the policy.

Summary

- We know about Open Access Data Policies .
- We mostly see it as another admin chore sent on our heads from above.
- We mostly think that it's not our business and we let other people to deal with that.

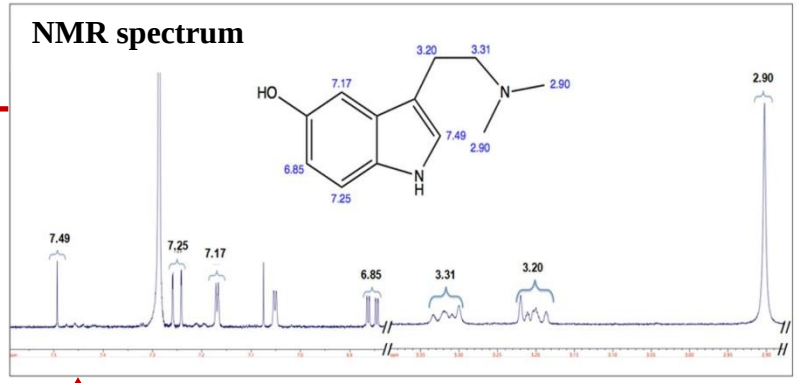
Let's try to see it from another perspective.

Golden Age of NMR Spectroscopy Workflow

Synthetic Lab

sample

NMR spectrum

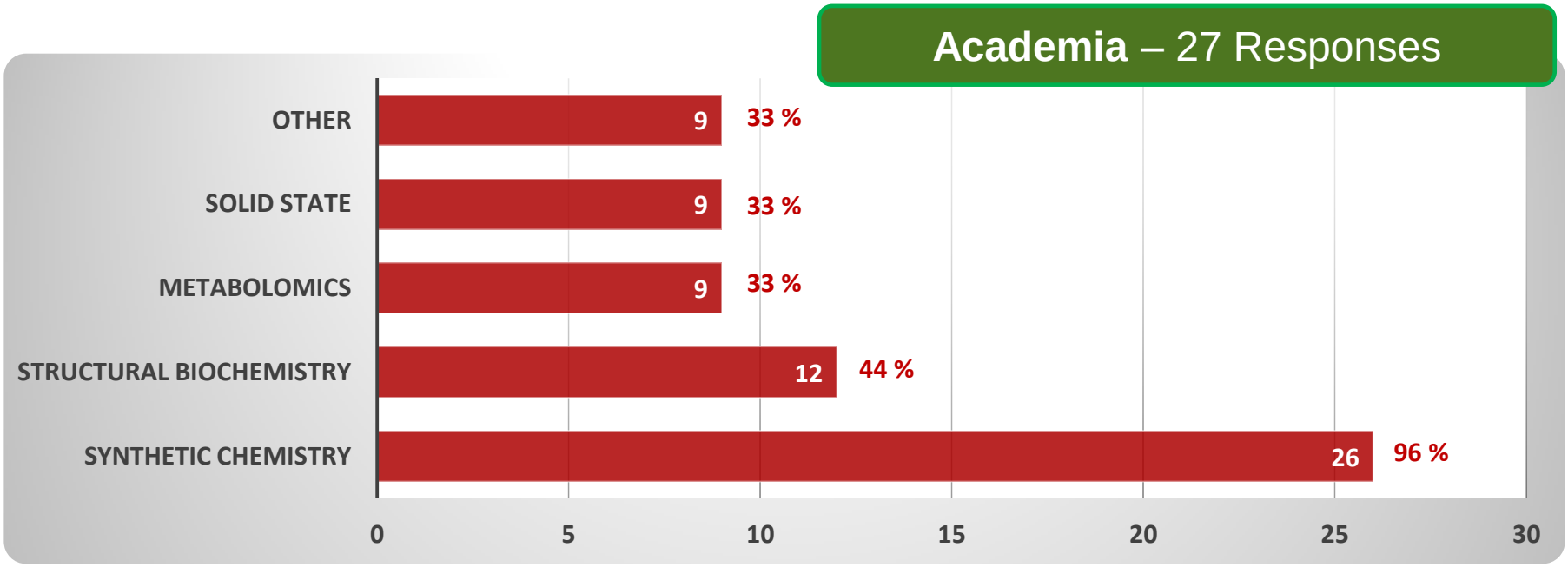


NMR Lab

Productivity – 10 to 20 experiments per day

Online survey – Current NMR dataflow

What type of chemistry does your lab support ?



Average NMR lab

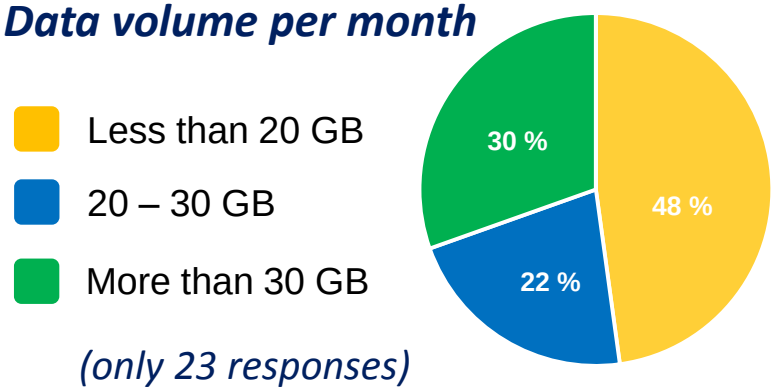
Number of instruments : 6

Number of sample changers : 4

Number of users : ~ 150

Number of experiments per month : ~ 4400

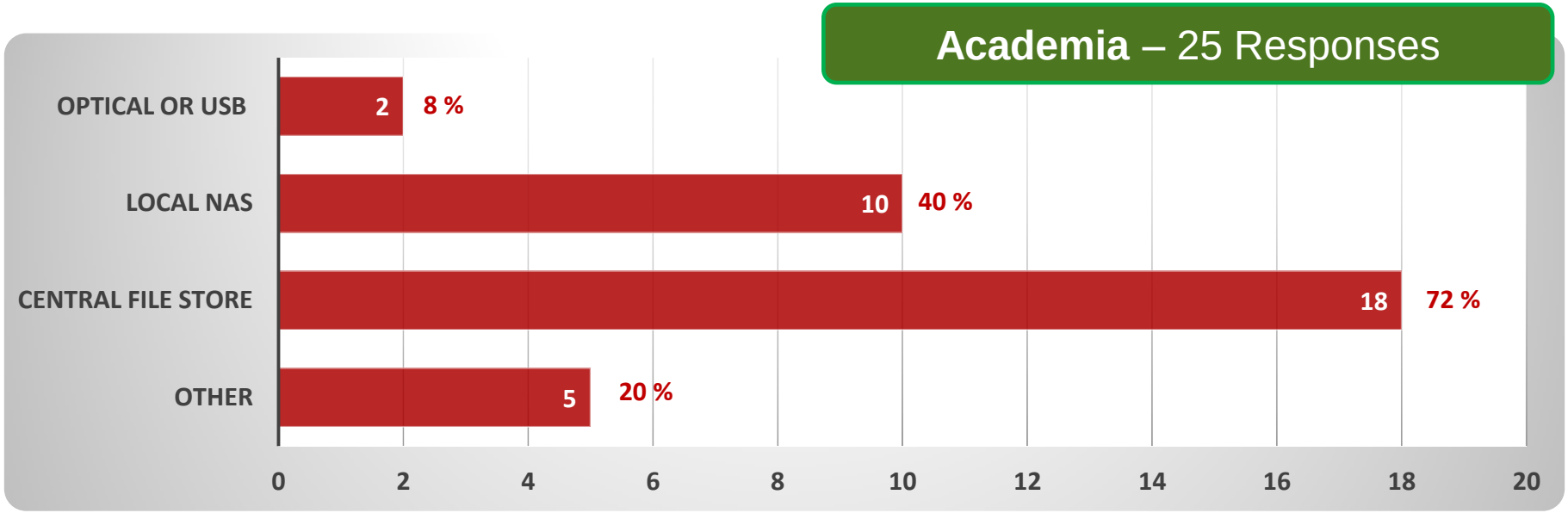
Data volume per month



(only 23 responses)

Online survey – Current NMR dataflow

Where do you archive NMR data in your lab?



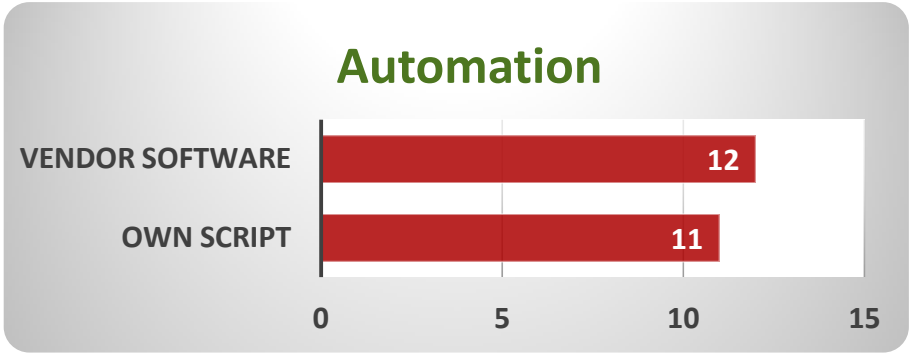
- Obvious trend towards Central File Storage systems
- Very likely impact of Open Access Data policies

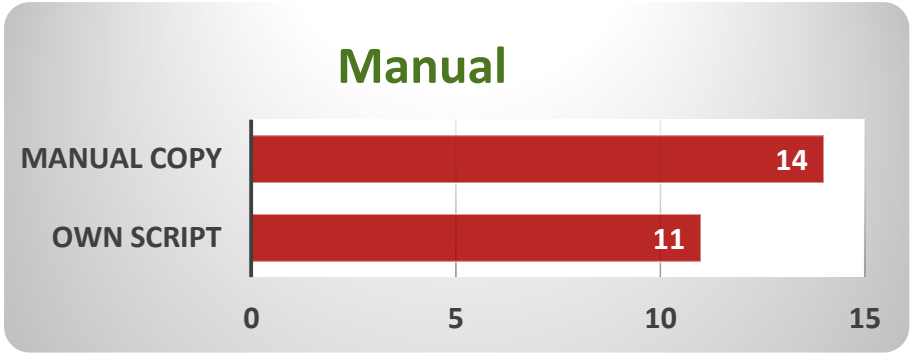
✓ **Positive**

More important is how we store and access the data

Online survey – Current NMR dataflow

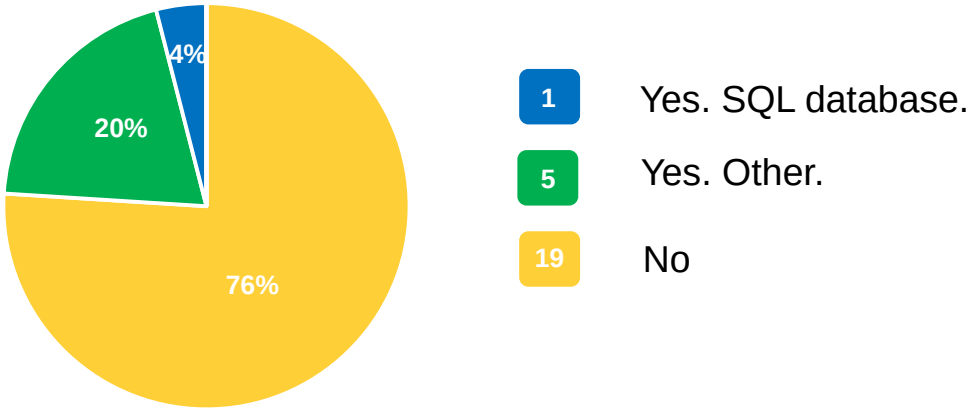
How do you archive NMR data in your lab?



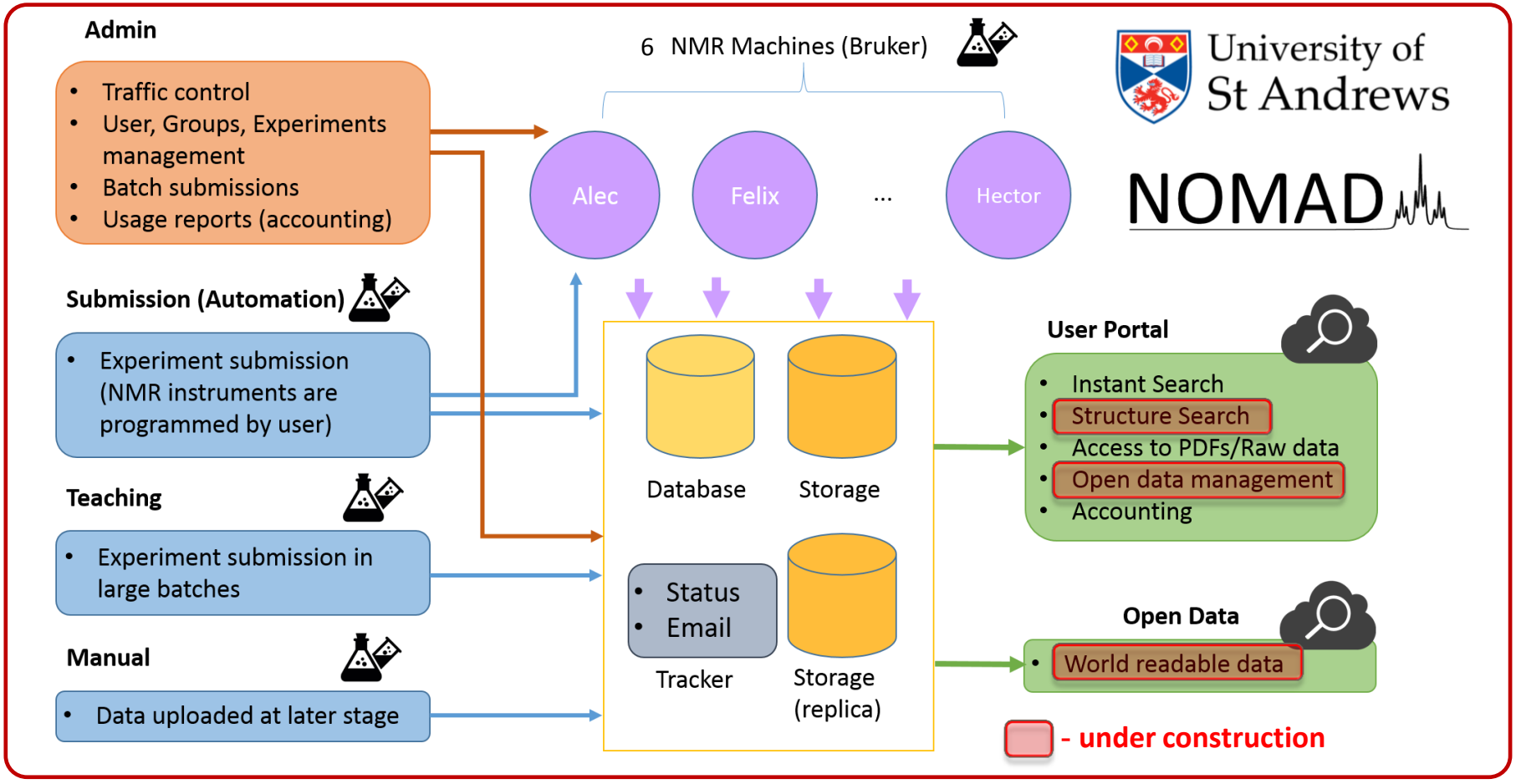


- Archiving script can be seen unsecure by people maintaining central file store.
- It seems that no one in UK academia uses commercial database for NMR spectra like ACD/Labs Spectrus DB, Mnova DB plugin or Magic Angle LOGS.

Does your NMR data store provide a search facility?



NOMAD – NMR Online Management And Datastore



- Web based "*cloud computing*" system that provides a resilient searchable data-store and centralised control of automation for Bruker NMR instruments.
- **Current version 1.2 (prototype)** – 6 Bruker instruments with sample changers, 260 research users, 350 users in teaching labs
- **Dataflow** – 6 000 to 10 000 Experiments (20 -30 GB) per month

NOMAD – Submission Portal

Always **REFRESH** this page before you decide/act, the list shown here is **00:00:34 (hh:mm:ss)** old and thus might be outdated

Submit Selected Jobs | BOOK NEW JOB | Cancel Selected Jobs

Holder	NMR Machine	Username	PIGroup	Solvent	Sample Identifier	ExpNo	Experiment	Parameters	Title	Print	Email	Date	Time	Status	Select
8	Alec	tl12	tl	CDCl3	09142015-8-tl-tl12-A	11	c13_deptq.c.and	ns,256	Demo		Y	2015-09-14	12:42	BOOKED	☐
8	Alec	tl12	tl	CDCl3	09142015-8-tl-tl12-A	10	proton.c.and		Demo		Y	2015-09-14	12:42	BOOKED	☐
Holder	**NMR Machine**	**Username**	**PIGroup**	**Solvent**	**Sample Identifier**	**ExpNo**	**Experiment**	**Parameters**	**Title**	**Print**	**Email**	**Date**	**Time**	**Status**	**Select**

Alec | Felix | Hector | Marcus | Noah | Robbin

Automation - Running - Busy until : Mon 13:05 - Day Experiments : 00:12 - Night Experiments : 00:00 - MAX ALLOWANCE : 00:15

Holder	NMR Machine	Username	PIGroup	Solvent	Sample Identifier	ExpNo	Experiment	Parameters	Title	Print	Email	Date	Time	Duration	Status	Holder	Select
11	Marcus	tb275	doh	CDCl3	09142015-11-doh-tb275-M	10	ghsqc-ed.a.and		tb150909		Y	2015-09-14	12:39	00:10:55	SUBMITTED	11	☐
15	Marcus	na10	doh	CDCl3	09142015-15-doh-na10-M	12	f19cpd.a.and		NAphenylethanol		Y	2015-09-14	12:35	00:01:54	RUNNING	15	☐
15	Marcus	na10	doh	CDCl3	09142015-15-doh-na10-M	11	pro19Fdec.a.and		NAphenylethanol		Y	2015-09-14	12:35	00:00:30	FINISHED	15	☐
15	Marcus	na10	doh	CDCl3	09142015-15-doh-na10-M	10	proton_WIDE.a.and		NAphenylethanol		Y	2015-09-14	12:35	00:00:51	FINISHED	15	☐
14	Marcus	rmnp	ads	CDCl3	09142015-14-ads-rmnp-M	10	proton_Quant.a.and		104 Aldehyde		Y	2015-09-14	12:08	00:05:58	FINISHED	14	☐

Username tl12

Password ••••••••

Name Tomas Lebl

PI Group tl

Available Machines Alec

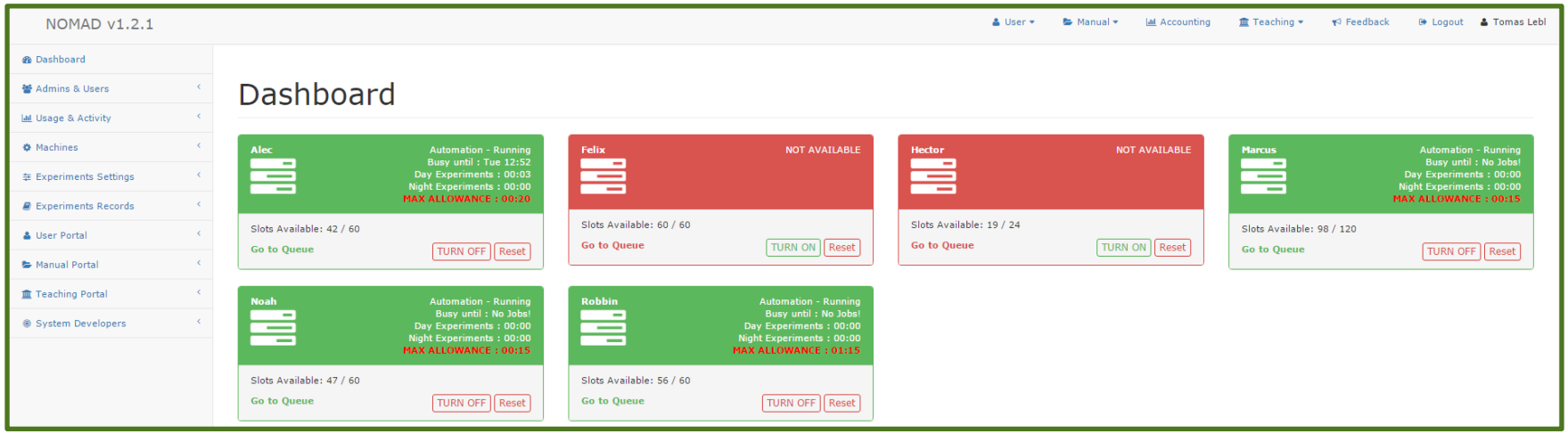
No of Samples 2

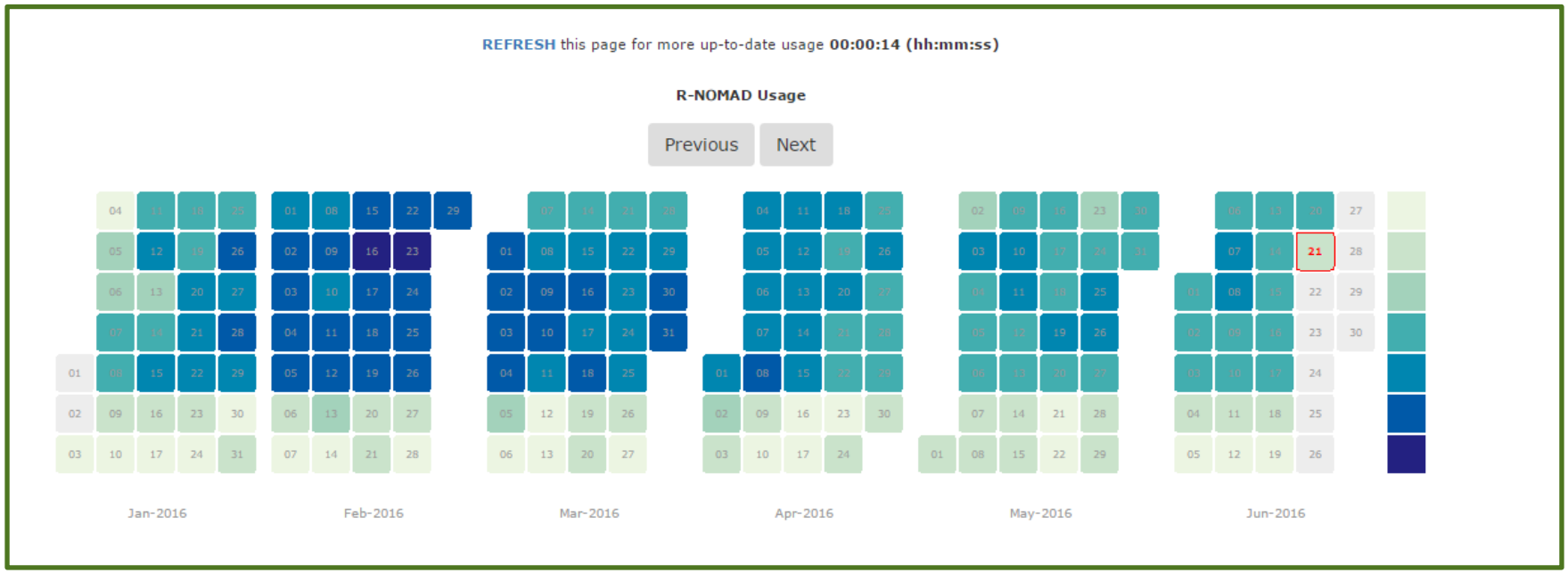
Holder	Sample Identifier	Solvent	Experiment	Parameters	Print	Email	Title	Duration	Total (D)	More
21	09142015-21-tl-tl12-A	CDCl3	proton.c.and		☐	☑	Demo 1	00:00:34	00:02:47	Add Del
			p31cpd.c.and		☐	☑	Demo 1	00:02:13		
22	09142015-22-tl-tl12-A	DMSO	proton.c.and	ns,512	☐	☑		00:48:33	00:48:33	Add

Continue | Start Over | Back To List

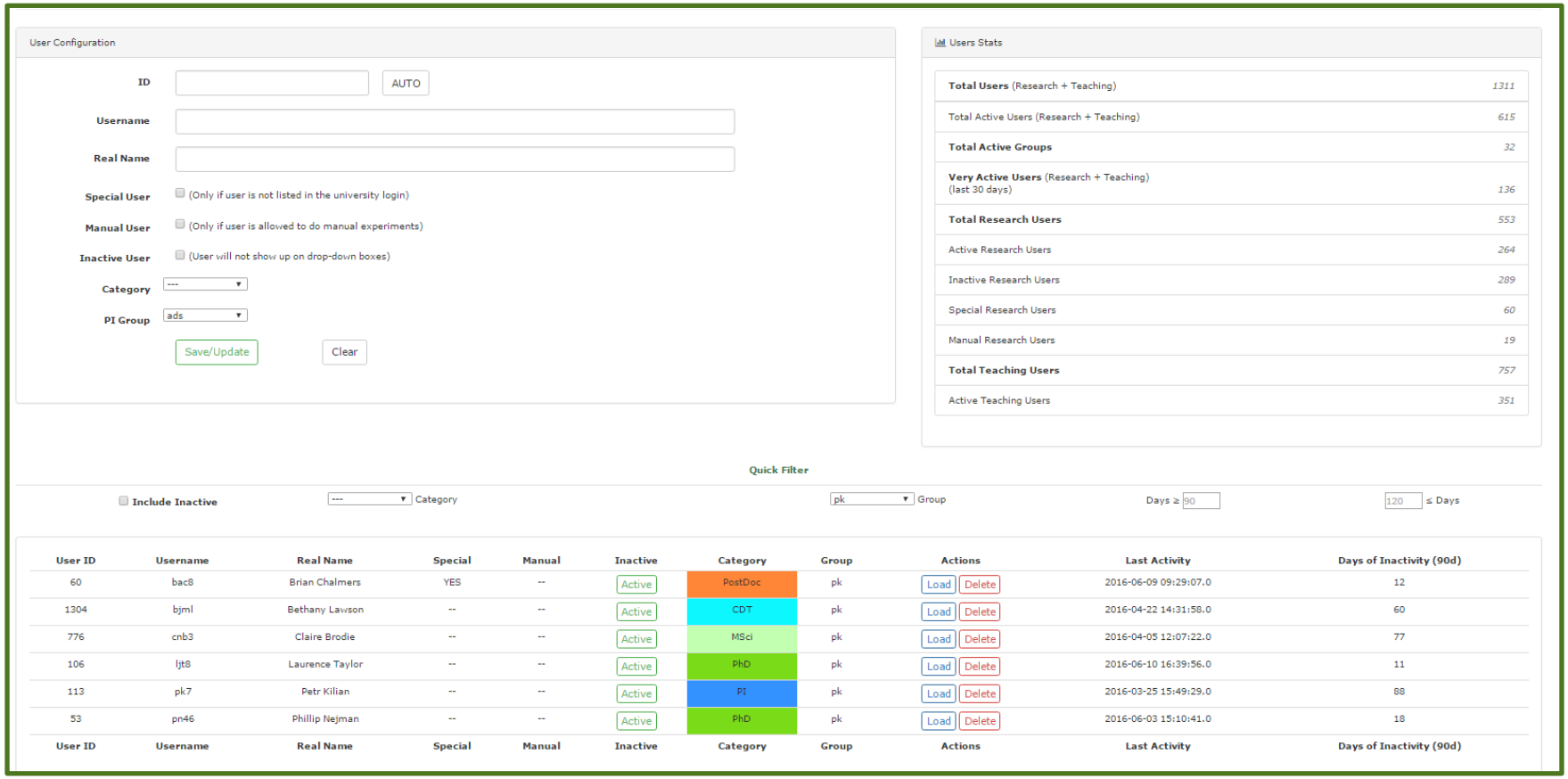
Holders exceeding (00:30:00) will be submitted as NIGHT experiments

NOMAD – Admin Portal





NOMAD – Admin Portal - user & experiment tables



User Configuration

ID [ ] AUTO

Username [ ]

Real Name [ ]

Special User ☐ (Only if user is not listed in the university login)

Manual User ☐ (Only if user is allowed to do manual experiments)

Inactive User ☐ (User will not show up on drop-down boxes)

Category [---]

PI Group [ads]

Save/Update Clear

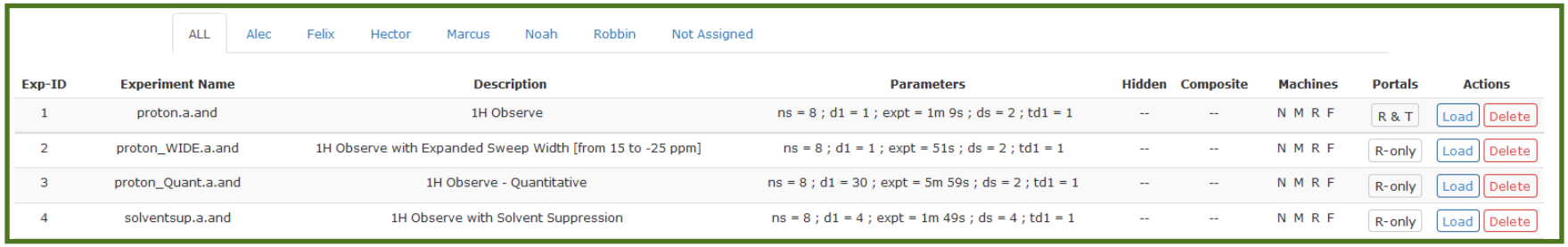
Users Stats

Total Users (Research + Teaching)	1311
Total Active Users (Research + Teaching)	615
Total Active Groups	32
Very Active Users (Research + Teaching) (last 30 days)	136
Total Research Users	553
Active Research Users	264
Inactive Research Users	289
Special Research Users	60
Manual Research Users	19
Total Teaching Users	757
Active Teaching Users	351

Quick Filter

☐ Include Inactive [---] Category [pk] Group Days ≥ [90] [120] ≤ Days

User ID	Username	Real Name	Special	Manual	Inactive	Category	Group	Actions	Last Activity	Days of Inactivity (90d)
60	bac8	Brian Chalmers	YES	--	Active	PostDoc	pk	Load Delete	2016-06-09 09:29:07.0	12
1304	bjml	Bethany Lawson	--	--	Active	CDT	pk	Load Delete	2016-04-22 14:31:58.0	60
776	cnb3	Claire Brodie	--	--	Active	MSci	pk	Load Delete	2016-04-05 12:07:22.0	77
106	ljt8	Laurence Taylor	--	--	Active	PhD	pk	Load Delete	2016-06-10 16:39:56.0	11
113	pk7	Petr Kilian	--	--	Active	PI	pk	Load Delete	2016-03-25 15:49:29.0	88
53	pn46	Phillip Nejman	--	--	Active	PhD	pk	Load Delete	2016-06-03 15:10:41.0	18
User ID	**Username**	**Real Name**	**Special**	**Manual**	**Inactive**	**Category**	**Group**	**Actions**	**Last Activity**	**Days of Inactivity (90d)**



ALL Alec Felix Hector Marcus Noah Robbin Not Assigned

Exp-ID	Experiment Name	Description	Parameters	Hidden	Composite	Machines	Portals	Actions
1	proton.a.and	1H Observe	ns = 8 ; d1 = 1 ; expt = 1m 9s ; ds = 2 ; td1 = 1	--	--	N M R F	R & T	Load Delete
2	proton_WIDE.a.and	1H Observe with Expanded Sweep Width [from 15 to -25 ppm]	ns = 8 ; d1 = 1 ; expt = 51s ; ds = 2 ; td1 = 1	--	--	N M R F	R-only	Load Delete
3	proton_Quant.a.and	1H Observe - Quantitative	ns = 8 ; d1 = 30 ; expt = 5m 59s ; ds = 2 ; td1 = 1	--	--	N M R F	R-only	Load Delete
4	solventsup.a.and	1H Observe with Solvent Suppression	ns = 8 ; d1 = 4 ; expt = 1m 49s ; ds = 4 ; td1 = 1	--	--	N M R F	R-only	Load Delete

NOMAD – Admin Portal - Accounting

Reports Search

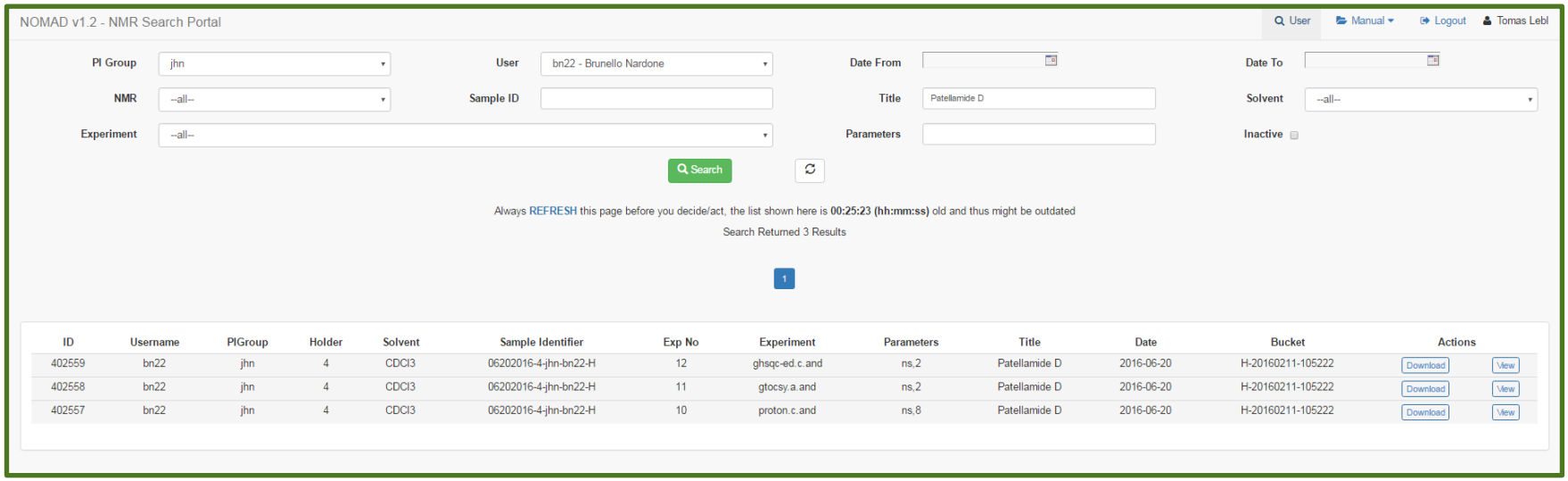
PI Group: dp Inactive Date From: 01/09/2015 Date To: 14/09/2015 View Duration Complete Clear

Storage Data (GB) Loading Status:

Complete

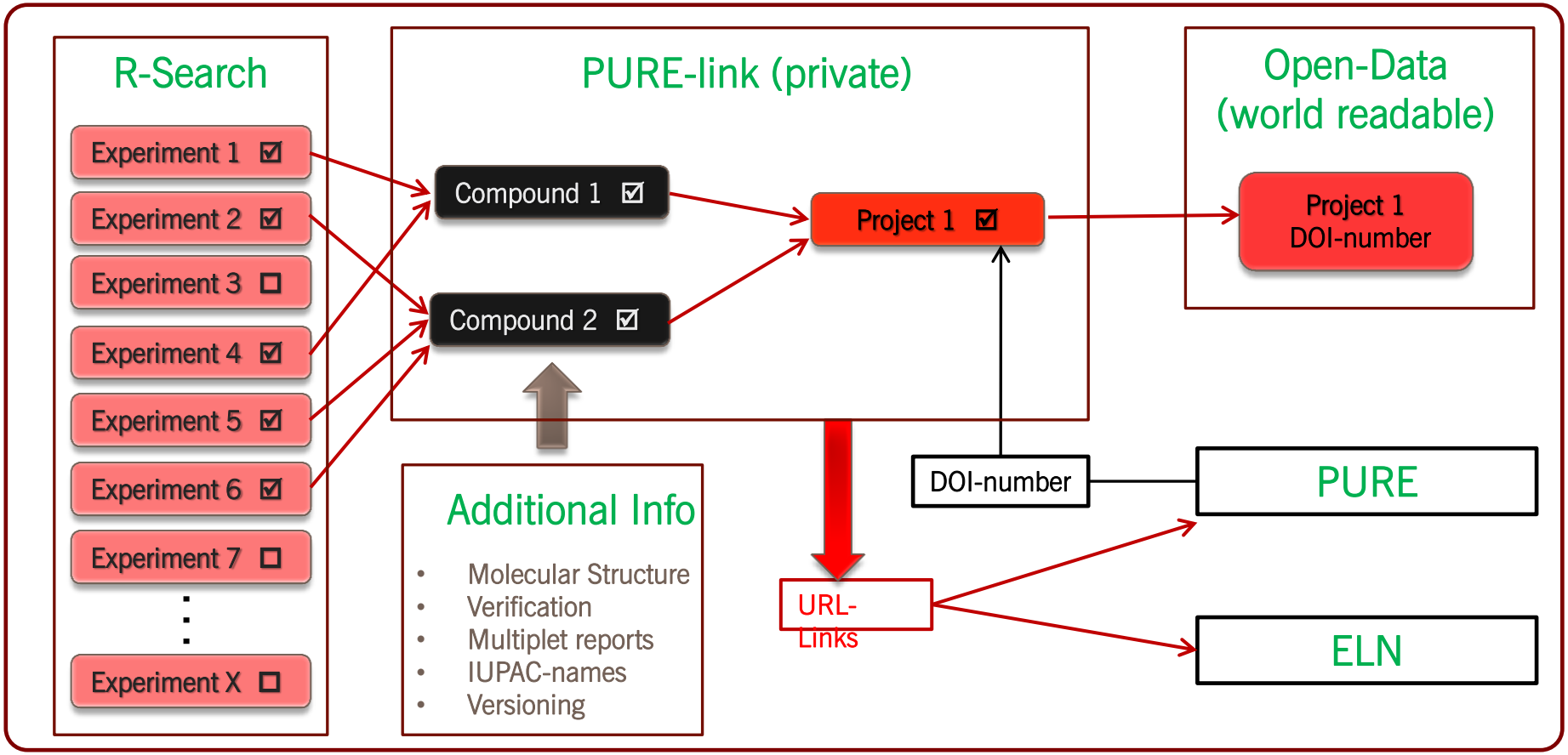
	Name	Alec			Felix			Hector			Marcus			Noah			Robbin			Total Cost
		ExpT	GB	Cost	ExpT	GB	Cost	ExpT	GB	Cost	ExpT	GB	Cost	ExpT	GB	Cost	ExpT	GB	Cost	
Auto		-	-	-	-	-	-	-	-	-	11:17:21	0.066	£46.80	08:55:11	0.057	£37.12	-	-	-	**£83.91**
Manual		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	dvb - David Van Brussel	N/A	-	-	N/A	-	-	N/A	-	-	N/A	0.066	£46.80	N/A	0.057	£37.12	N/A	-	-	**£83.91**
Auto		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Manual		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	dp12 - Douglas Philp	N/A	-	-	N/A	-	-	N/A	-	-	N/A	-	-	N/A	-	-	N/A	-	-	-
Auto		00:40:40	0.009	£6.00	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	**£6.00**
Manual		-	-	-	2h	0.066	£12.65	-	-	-	-	-	-	-	-	-	-	-	-	**£12.65**
	jr74 - Josh Richards	N/A	0.009	£6.00	N/A	0.066	£12.65	N/A	-	-	N/A	-	-	N/A	-	-	N/A	-	-	**£18.65**
Auto		00:58:23	0.015	£8.65	00:31:52	0.006	£3.08	16:50:46	0.100	£170.97	00:12:57	0.002	£0.92	00:47:42	0.007	£3.36	00:55:56	0.006	£2.48	**£189.45**
Manual		-	-	-	-	-	-	1d - 10h	0.035	£220.88	-	-	-	-	-	-	-	-	-	**£220.88**
	lm216 - Leonardo Maugeri	N/A	0.015	£8.65	N/A	0.006	£3.08	N/A	0.135	£391.84	N/A	0.002	£0.92	N/A	0.007	£3.36	N/A	0.006	£2.48	**£410.33**
Auto		00:33:02	0.004	£4.78	00:07:06	0.002	£0.70	-	-	-	-	-	-	-	-	-	-	-	-	**£5.48**
Manual		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	tk235 - Tamara Kosikova	N/A	0.004	£4.78	N/A	0.002	£0.70	N/A	-	-	N/A	-	-	N/A	-	-	N/A	-	-	**£5.48**
Auto		02:12:05	0.028	£19.42	00:38:58	0.008	£3.78	16:50:46	0.100	£170.97	11:30:18	0.068	£47.71	09:42:53	0.065	£40.48	00:55:56	0.006	£2.48	**£284.84**
Manual		-	-	-	2h	0.066	£12.65	1d - 10h	0.035	£220.88	-	-	-	-	-	-	-	-	-	**£233.53**
	TOTAL	N/A	0.028	£19.42	N/A	0.074	£16.43	N/A	0.135	£391.84	N/A	0.068	£47.71	N/A	0.065	£40.48	N/A	0.006	£2.48	**£518.37**
	Name	**Alec**			**Felix**			**Hector**			**Marcus**			**Noah**			**Robbin**			**Total Cost**
		ExpT	**GB**	**Cost**	**ExpT**	**GB**	**Cost**	**ExpT**	**GB**	**Cost**	**ExpT**	**GB**	**Cost**	**ExpT**	**GB**	**Cost**	**ExpT**	**GB**	**Cost**	

NOMAD – User Search Portal



NOMAD – PURE link & Open-Data

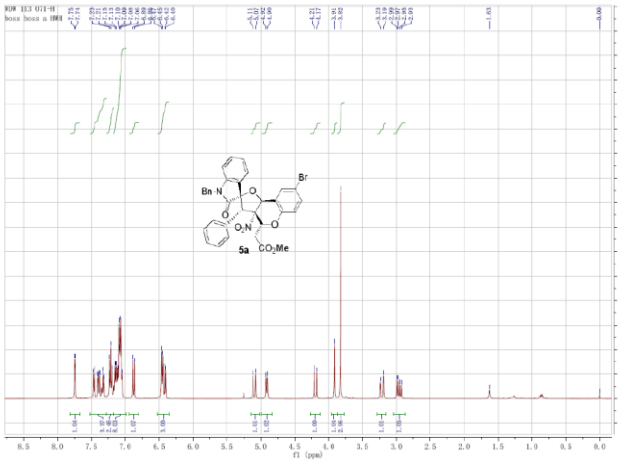
- NOMAD stores the data securely and provides instant search facility but the data are not publicly available.
- Users have to download selected data and upload them to PURE (Research Information Management System).
- It's not clear what should be uploaded to PURE and capacity is limited.

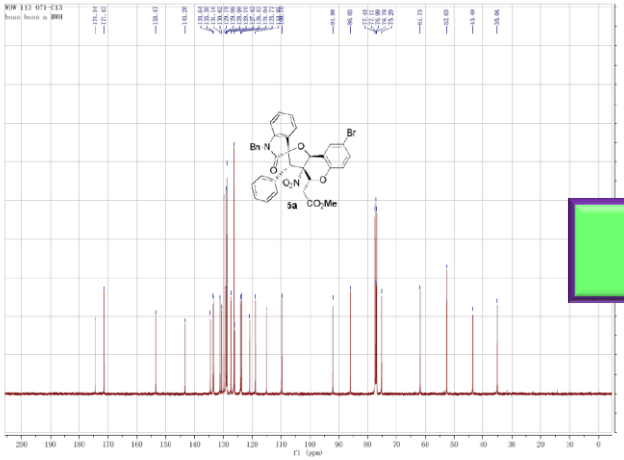


Open Access Data Policies – Good or Bad or Ugly

- ***Nowadays NMR spectroscopy produces vast quantity of data.***
  - Is that data sufficiently utilised and shared?

    Pictures of ^{1}H and ^{13}C spectra in Supporting Information





I don't think so!

- Why academia does not use commercial NMR databases?
- Why NMR spectroscopy does not have an equivalent of Protein Data Bank or Cambridge Crystallographic Data Centre?

Likely there has been no driving force.

Could Open Access Data Policies become that driving force?

Maybe

NOMAD - Current Status and Future Plan

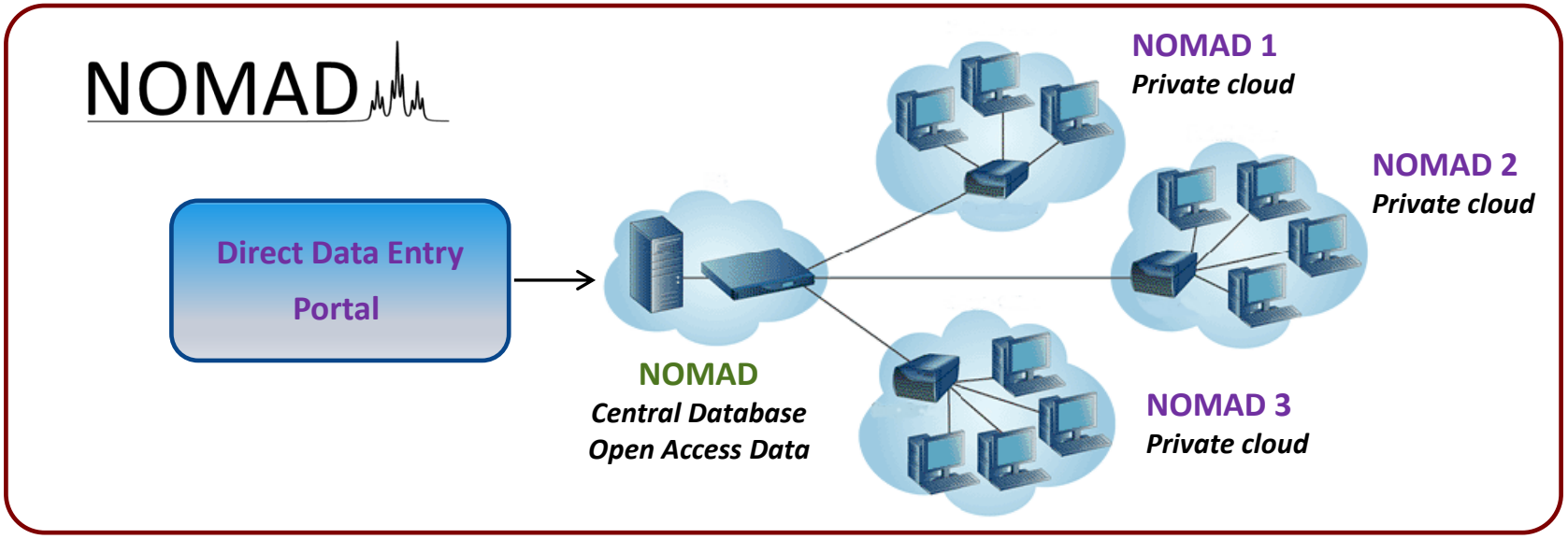
Current status

- Team of 6 people working part time
- We have setup infrastructure and processes for software development and production
- IAA funding available until March 2017

Short term future plan

- “PURE link” add-on function and “Open Data Portal”
- NOMAD 2.0 – modular architecture, separate database and control function

Long term future plan



Discussion

- Do you like the concept of NOMAD system? Any comments ?
- If yes how to propel further development and distribution?
    a) University of St Andrews spin-off
    b) EPSRC or other funding